Far East Journal of Mathematical Sciences (FJMS)

http://www.pphmj.com
http://dx.doi.org/10.17654/MS102040779
Volume 102, Number 4, 2017, Pages 779-791 **ISSN: 0972-0871**

EFFECTIVE TECHNIQUES FOR ESTIMATING THE PARAMETERS OF GRUBBS MODEL WITH APPLICATIONS

Almousa Amani, Riad Mahmoud and Mansour Hanadi

Institute of Statistical Studies and Research
Cairo University
Egypt

Abstract

In this paper, three resampling techniques are considered, namely, bootstrap, jackknife, and jackknife after bootstrap. The main objective is to study the performance of these techniques in the maximum likelihood estimation of the parameters for Grubbs model. Also, the performance of these techniques is discussed in the detection of the influential observations using local influence method under different perturbation schemes for Grubbs model. The performance is illustrated through an application using real data set. Our results provide resampling techniques offer better fit, protection against outliers and more precise inferences than the traditional methods.

1. Introduction

The problem of comparing measurement devices which vary in price, time spent to measure and other features, such as efficiency, has been of growing interest in many scientific applications. Grubbs measurement error

Received: April 5, 2017; Accepted: June 11, 2017

2010 Mathematics Subject Classification: 62-XX.

Keywords and phrases: expectation maximization, Grubbs model, jackknife after bootstrap, scale mixture of normal distributions.

model was introduced in [14]. This model is typically used in comparison studies to assess the relative agreement between two or more measuring devices (or instruments) that are used to measure the same quantity of interest. [15] studied the normal Grubbs (N) model, noting the well-known lack of robustness of least-square estimates against outlying observations. To overcome this deficiency, a general class of scale mixture of normal Grubbs model (SMN) was proposed in [23]. Properties of the SMN distributions, such as student-t (T), slash normal (SL) and contaminated normal (CN) may found in [16, 23]. In an asymmetric setting, [22] proposed the skew normal Grubbs (SN) model and showed advantages of using asymmetric distributions for obtaining accurate robust estimates. Three resampling techniques will be considered in this paper. Bootstrap was proposed in [9], jackknife in [10] and jackknife after bootstrap (JaB) in [11]. Bootstrap and jackknife resampling techniques were considered in [12, 24] and they obtained the bootstrap and jackknife estimates in linear regression model. Also, both resampling were used in [1] to estimate the sampling distribution of the parameter estimates in linear regression model. JaB was used in [18] to determine the cut-off values for various diagnostic measures in linear regression under non-normal errors and small samples. JaB was used in [2] in count regression model to assess the error in the bootstrap estimate parameters. Also, JaB was used in [5] to detect influential observations for binary logistic regression model. The primary objective of this paper is to illustrate the performances of resampling techniques such as bootstrap, jackknife and JaB in the estimation of the parameters for N, SN and SMN models, respectively, in addition, to detect the influential observations using the local influence method under different perturbation schemes following [15, 21, 23]. This paper is organized as follows: Section 2 describes the Grubbs model. Sections 3-5 discuss the performance of the resampling techniques in the estimation of the parameters, in the detection of the outlying observations and in the detection of the influential observations, respectively. Section 6 illustrates the performance of the resampling techniques through application using real data. Section 7 contains conclusions.

2. The Grubbs Model

Let x_i be the unobserved value corresponding to unit i, and y_{ij} be the measured value obtained with the instrument j in unit i, $i = 1, \ldots, n$ and $j = 1, \ldots, p$, with $p \geq 2$. The normal Grubbs model can be defined as [22]

$$y_i = \alpha + 1_p x_i + \varepsilon_i, \tag{1}$$

where $\alpha = (0, \alpha_2, \ldots, \alpha_p)^T$, and $1_p = (1, \ldots, 1)^T$ are $p \times 1$ vectors; $y_i = (y_{i1}, \ldots, y_{ip})^T$ and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{ip})^T$ (the error vector) are $p \times 1$ random vectors independent with $\varepsilon_i \overset{iid}{\sim} N_p(0, D(\phi))$ and $x_i \overset{iid}{\sim} N_1(\mu_x, \phi_x)$, where $D(\phi) = (\phi_1, \ldots, \phi_p)^T$.

3. Resampling Algorithmic Approach for Estimation of Parameters for Grubbs Model

This section is devoted to study the performance of resampling techniques such as bootstrap, jackknife and JaB in the estimation of the parameters for Grubbs models. Bootstrap and jackknife estimates of the parameters and the relevant standard error (SE) for Grubbs model can be obtained following the bootstrap and the jackknife algorithms described in [24], but the estimation of the parameters is to be conducted using the expectation maximization (EM) algorithm [7] for N and SN models, following [15, 22]. Using expectation conditional maximization (ECM) algorithm [20] for SMN models (T, SL, CN) following [23], jackknife after bootstrap (JaB) estimates of the parameters and the relevant standard error for each estimate are obtained in Section 5 of [2].

4. Resampling Algorithmic Approach for Detection of the Outlying Observations for Grubbs Model

To study the performance of resampling techniques in the detection of the outlying observations, we can use the Mahalanobis distance as a diagnostic measure following [15, 23]. Jackknife and JaB techniques are considered in this section. For jackknife technique, we can detect the outlying observations, by removing the *i*th data point from the original data set, and compute the Mahalanobis distance for $(n-1)$ observations, then a point is flagged as outlier if its distance value exceeds the cut-off value. The cut-off value depends on the distribution of the Mahalanobis distance for N, SN, T, SL and CN distributions [16, 21]. After determining the outliers for each jackknife resample, calculate the percentage of data sets among all *n* reduced data sets in which each data point is flagged. This overall percentage will typically be a large indicator for this point to be an outlier, following [19, Section 2].

For JaB technique, we can find the appropriate JaB influence cut-offs for a Mahalanobis distance to detect the outlying observations for Grubbs model following algorithm in Section 2 of [18].

5. Resampling Algorithmic Approach for Detection of the Influential Observations for Grubbs Model

The local influence approaches [6, 26] were applied in [15] for N, [21] for SN and [23] for SMN models. Now we use these approaches for studying the performance of jackknife and JaB in the detection of the influential observations for each distribution.

For jackknife technique, we can detect the influential observations by the same method as described in the above section but the diagnostic measure is the conformal normal curvature

$$M(0) = B_{f_Q, h_j} = C_{f_Q, h_j} / trace(2\Delta^T \{-\ddot{Q}\} \Delta), \quad j = 1, \ldots, q. \tag{2}$$

The cut-off value $= \overline{M}(0) + c^* SD(M(0))$, $\overline{M}(0) = 1/q$, q is the dimension of the perturbation vector, c^* is a selected constant, $SD(M(0))$ is the standard deviation of $M(0)$. The delta matrix Δ for each perturbation and for each distribution can find it in [15] for N, [21] for SN and [23] for SMN models. For JaB technique, we can find an appropriate influence of cut-offs for a conformal curvature normal measure to detect the influential observations for Grubbs models, following [18, Section 2].

6. Application

A real example is presented to illustrate the performance of resampling techniques in the estimation of the parameters for Grubbs model and influence diagnostics. The real data considered are the Barnett data sets [3]. The Barnett data sets described as two instruments were used for measuring the vital capacity of the human lung that operated by skilled and unskilled operators that were compared on a common group of 72 patients. The following four instruments were compared: Instrument 1: Standard instrument and skilled operator; Instrument 2: Standard instrument and unskilled operator; Instrument 3: New instrument and skilled operator and Instrument 4: New instrument and unskilled operator. The analysis conducted using R version 3.3.1. The number of replications (B) will be equal to 100, 1000. The convergence criterion

$$\max_{j=1,\ldots,n_p} \left| \frac{\hat{\theta}^{(k+1)} - \hat{\theta}(k)}{\hat{\theta}^{(k)}} \right| \leq \delta,$$

where n is the dimension of θ, δ is very small number, say 10^{-6}, is used for the EM and ECM algorithms.

Bootstrap, jackknife and JaB estimates and the relevant standard error of these estimates for N, SN, T, SL and CN models are obtained, corresponding the number of replications $B = 100, 1000$, as can be seen in Table 1. Noting that JaB estimates and bootstrap estimates are very close for each replication for all distributions. Also, the jackknife estimates and the original estimates for all distributions are very close. Bootstrap and JaB standard error (SE) are not comparable because there are different scales when $B = 100, 1000.$ But the (SE) values of jackknife estimates are less than bootstrap and JaB SE for all distributions. JaB estimates are not shown in Table 1 because it is identical with bootstrap estimates when $B = 100, 1000.$

Table 1. The ML estimates for N, SN, T, SL, CN models for jackknife, bootstrap when $B = 100, 1000$

Resampling technique	Parameter	CN	T	SL	N	SN
Original sample	μ_x	21.02	20.93	21.02	22.46	12.15
	φ_x	65.03	45.83	86.69	62.91	168.98
	φ_1	5.00	3.03	7.14	4.99	5.06
	φ_2	1.41	0.82	2.09	1.41	1.25
	φ_3	4.37	2.83	5.86	4.38	4.53
	φ_4	4.62	3.19	6.02	4.63	4.76
	α_2	–0.70	–0.63	0.28	–0.70	–0.70
	α_3	–0.98	–0.92	0.12	–0.97	–0.98
	α_4	–1.44	–1.22	–0.47	–1.44	–1.44
	λ_x					5.68
Jackknife	μ_x	21.02	20.939	21.02	22.46	12.14
	φ_x	65.03	45.83	86.69	52.89	169.26
	φ_1	5.00	3.03	7.14	4.99	5.06
	φ_2	1.42	0.82	2.09	1.41	1.25
	φ_3	4.37	2.83	5.86	4.38	4.53
	φ_4	4.62	3.19	6.02	4.63	4.76
	α_2	–0.70	–0.63	0.28	–0.70	–0.70
	α_3	–0.98	–0.92	0.12	–0.97	–0.97
	α_4	–1.44	–1.22	–0.47	–1.44	–1.44
	λ_x					5.76

Bootstrap (100)	μ_x	21.029	20.909	20.995	22.437	12.805
	φ_x	65.265	45.021	88.661	61.671	153.537
	φ_1	4.904	2.906	6.879	4.962	5.074
	φ_2	1.415	0.791	2.231	1.376	1.154
	φ_3	4.345	2.752	5.827	4.345	4.403
	φ_4	4.703	3.111	5.952	4.745	4.652
	α_2	–0.734	–0.609	–0.709	–0.729	–0.750
	α_3	–0.988	–0.897	–0.841	–1.063	–1.001
	α_4	–1.405	–1.198	–1.425	–1.501	–1.494
	λ_x					6.732
Bootstrap (1000)	μ_x	21.009	20.925	21.019	22.495	12.891
	φ_x	64.916	45.572	86.364	62.132	157.170
	φ_1	4.966	3.069	6.997	4.950	5.056
	φ_2	1.416	0.832	2.045	1.387	1.236
	φ_3	4.344	2.770	5.856	4.365	4.376
	φ_4	4.613	3.118	6.046	4.609	4.742
	α_2	–0.698	–0.632	–0.718	–0.713	–0.734
	α_3	–0.951	–0.921	–0.864	–0.971	–0.967
	α_4	–1.426	–1.233	–1.459	–1.445	–1.442
	λ_x					6.291

Bootstrap, jackknife and JaB absolute bias of the estimates are computed, as seen in Table 2. Noting that bootstrap absolute bias when $B = 1000$ less than bootstrap absolute bias when $B = 100$ for all distributions, also the same result for JaB. The relative efficiencies $RE1 = SE(\hat{\theta}_{boot})/SE(\hat{\theta}_{jack})$ and $RE2 = SE(\hat{\theta}_{jab})/SE(\hat{\theta}_{jack})$ are computed for each resampling technique. We can conclude that the RE2 for JaB when $B = 1000$ is larger than the RE2 for JaB when $B = 100$ for all estimates except ϕ_x for N, SN cases, but for SL, CN, T, the estimates μ_x, $\phi_x \alpha_3$, α_4 are to be expected. The same result for bootstrap indicates that the resampling techniques lead to efficient results. The results are summarized in Table 2 for N and CN only.

Table 2. Bias estimate and the relative efficiency for all models for each resampling technique, when $B = 100, 1000$

		Absolute bias			Relative efficiency			
Parameter	Distribution	Jackknife	Bootstrap (100)	Bootstrap (1000)	Bootstrap (100)	Bootstrap (1000)	JaB (100)	JaB (1000)
μ_x	N	1.46	1.44	1.495	1.69	1.23	5.23	1.44
φ_x		5.11	6.329	5.868	72	6.43	38.93	15.64
φ_1		54.01	54.038	54.05	78.53	57.68	212.1	149.9
φ_2		60.59	60.624	60.613	64.4	64.1	485.6	412.4
φ_3		74.62	74.655	74.635	67.26	68.1	166.1	139.7
φ_4		68.37	68.255	68.391	68.26	64.53	157.4	133.7
α_2		0	0.029	0.013	18.53	17.65	47.71	44.82
α_3		0.07	0.163	0.071	21.79	21.42	29.63	28.32
α_4		0.06	0.001	0.055	22.68	21.53	33.74	28.89
μ_x	CN	0.02	0.029	0.009	8.514	8.59	6.62	5.91
φ_x		2.97	2.735	3.084	68.92	8	5.59	1.66
φ_1		54	54.096	54.034	9.07	10.17	10.20	14.3
φ_2		60.58	60.585	60.584	8.77	8.563	10.98	12.66
φ_3		74.63	74.655	74.656	10.98	10.497	10.72	12.46
φ_4		68.38	68.297	68.387	10.74	10.629	13.96	12.86
α_2		0	0.034	0.002	8.44	8.523	10.04	9.44
α_3		0.08	0.088	0.051	8.86	8.768	7.12	6.75
α_4		0.06	0.095	0.074	8.49	8.690	6.45	6.45

The values of the information criteria, the Akaike information criterion (AIC), the Schwarz Bayesian information criterion (BIC) and the Hannan-Quinn criterion (HQ) are used for comparing between distributions for each resampling technique, following [15, 21]. We can note that the values of the information criteria decrease when the number of replication increases, indicating that the distributions with jackknife, bootstrap and JaB offer better fit corresponding distributions with the original sample, while jackknife technique is the best technique. Also, bootstrap and JaB techniques outperform when the number of replications increases. The results are shown in Table 3 for N and SL distributions.

Table 3. Values of the information criteria for (N, SL) models for each resampling technique when $B = 100$, 1000

Distribution	B	Resampling techniques	Log likelihood	AIC	BIC	HQ
N		Original sample	–747.789	756.789	773.273	763.395
		Jackknife	–737.329	746.329	762.813	752.935
	100	Bootstrap	–742.769	751.769	768.253	758.375
		JaB	–742.769	751.769	768.253	758.375
	1000	Bootstrap	–742.701	751.701	768.184	758.307
		JaB	–742.701	751.701	768.184	758.307
SL		Original sample	–658.057	667.057	683.541	673.663
		Jackknife	–648.849	657.849	674.333	664.455
	100	Bootstrap	–654.876	663.876	680.359	670.482
		JaB	–654.876	663.876	680.359	570.482
	1000	Bootstrap	–653.603	662.603	679.087	669.209
		JaB	–653.603	662.603	679.087	669.209

In order to detect outlying observations, consider the Mahalanobis distance (d_i), adopting the cut-off lines which correspond to the quantile 0.95. Noting that the JaB technique flagged fewer outlying observations than original and jackknife as seen in Table 4, indicating that JaB technique is more robust to outlying observations for all distributions.

Table 4. The outlying observations for N, SN, CN models for jackknife and JaB resampling techniques when $B = 100$

Resampling technique	Distribution	Cut-off	Outliers
Original sample	N	9.48773	1, 7, 36, 45, 49, 52, 62
	SN	9.48773	1, 7, 36, 49, 62, 72
	CN	9.3999	1, 7, 36, 45, 49, 52, 62
Jackknife	N	9.48773	1, 7, 36, 44, 48, 51, 61
	SN	9.48773	5, 7, 36, 47, 48, 70, 71
	CN	9.3999	1, 7, 44, 48, 61
JaB	N	10.62683	7, 36, 49, 62
	SN	10.94074	7, 36, 49, 62
	CN	10.9077	7, 36, 49, 62

Table 5. The influential observations for N, T, SL, CN models under case weight and measurement for a particular instrument perturbation schemes and for jackknife and JaB resampling techniques when $B = 100$

	Perturbation scheme	Case weight		Perturbation for a particular instrument	
Distribution	Resampling technique	Cut-off	Influential observations	Cut-off	Influential observations
N	Original sample	0.0335	-	0.0452	5, 44, 45, 59
	Jackknife	0.0229	-	0.0452	5, 43, 44, 58
	JaB	0.0353	-	0.0479	5, 44, 45, 59
CN	Original sample	0.0441	1, 5, 27, 36, 44, 45, 59	0.0441	-
	Jackknife	0.0843	5, 44, 58	0.0448	-
	JaB	0.0849	5, 45, 59	0.0384	-
SL	Original sample	0.0911	1, 5, 45, 59	0.0552	3, 12, 16, 20, 32
	Jackknife	0.0912	1, 5, 44, 58	0.0596	3, 12, 27
	JaB	0.0991	5, 45, 59	0.1259	-
T	Original sample	0.0603	5, 44, 45, 56, 59	0.0361	18, 41, 51, 58
	Jackknife	0.0606	5, 44, 55, 58	0.0362	18, 40, 50, 57
	JaB	0.0653	44, 45, 59	0.0373	18, 41, 51, 58

To identify the influential observations using the local influence approach, the values of the conformal normal curvature B_i for each distribution for jackknife and JaB techniques are obtained under different perturbation schemes when $B = 100.$ We note from Table 5 that the observations 5, 45, 59 are the popular influential under case weight perturbation for all distributions except normal distribution. We can noted that JaB technique flagged fewer influential observations than original and jackknife for SN, T, SL, CN distributions under case weight perturbation. Under joint response perturbation and multiplicative bias perturbation, there is no influential observation appearing for all distributions.

7. Conclusions

The main conclusion is that the use of resampling techniques offers better fits and protection against outliers and more precise inferences than the traditional technique. The nature of the JaB method needs much computation especially for larger sample sizes. We can apply the resampling techniques for asymmetric version of SMN distributions proposed in [25] to accommodate skewness and heavy-tailedness simultaneously for Grubbs model called *scale mixtures of skew-normal* (*SMSN*) *distributions* and we can use the expectation of conditional maximization either (ECME) algorithm [17] to estimate the parameters for SMSN models following [8] and [13] as interesting topics for future research.

References

[1] Z. Y. Algamal and K. B. Rasheed, Resampling in linear regression model using jackknife and bootstrap, Iraqi Journal of Statistical Science 18 (2010), 59-73.

[2] Z. Y. Algamal and K. B. Rasheed, Re-sampling techniques in count data regression models, Iraqi Journal of Statistical Sciences 22 (2012), 15-25.

[3] V. D. Barnett, Simultaneous pairwise linear structural relationships, Biometrics 25 (1969), 129-142.

[4] U. Beyaztas and A. Alin, Jackknife after bootstrap method for detection of influential observations in linear regression models, Comm. Statist. Simulation Comput. 42 (2013), 1256-1267.

[5] U. Beyaztas and A. Alin, Jackknife-after-bootstrap as logistic regression diagnostic tool, Comm. Statist. Simulation Comput. 43 (2014), 2047-2060.

[6] R. D. Cook, Assessment of local influence, J. R. Stat. Soc. Ser. B Stat. Methodol. 48 (1986), 133-169.

[7] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1977), 1-38.

[8] C. da Silva Ferreira, H. Bolfarine and V. H. Lachos, Skew scale mixtures of normal distributions: properties and estimation, Stat. Methodol. 8 (2011), 154-171.

[9] B. Efron, Bootstrap Methods: Another Look at the Jackknife in Breakthroughs in Statistics, Springer, Vol. 7, 1979, pp. 1-26.

[10] B. Efron, The Jackknife, the Bootstrap and other Resampling Plans, SIAM, Vol. 38, 1982.

[11] B. Efron, Jackknife-after-bootstrap standard errors and influence functions, J. R. Stat. Soc. Ser. B Stat. Methodol. 54 (1992), 83-127.

[12] B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap, CRC Press, 1993.

[13] C. S. Ferreira, V. H. Lachos and H. Bolfarine, Inference and diagnostics in skew scale mixtures of normal regression models, J. Stat. Comput. Simul. 85 (2015), 517-537.

[14] F. E. Grubbs, On estimating precision of measuring instruments and product variability, J. Amer. Statist. Assoc. 43 (1948), 243-264.

[15] V. H. Lachos, F. Vilca and M. Galea, Influence diagnostics for the Grubbs' model, Statist. Papers 48 (2007), 419-436.

[16] K. Lange and J. S. Sinsheimer, Normal/independent distributions and their applications in robust regression, J. Comput. Graph. Statist. 2 (1993), 175-198.

[17] C. Liu and D. B. Rubin, The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence, Biometrika 81 (1994), 633-648.

[18] M. A. Martin and S. Roberts, Jackknife-after-bootstrap regression influence diagnostics, J. Nonparametr. Stat. 22 (2010), 257-269.

[19] M. A. Martin, S. Roberts and L. Zheng, Delete-2 and Delete-3 jackknife procedures for unmasking in regression, Aust. N. Z. J. Stat. 52 (2010), 45-60.

[20] X. L. Meng and D. B. Rubin, Maximum likelihood estimation via the ECM algorithm: a general framework, Biometrika 80 (1993), 267-278.

[21] L. C. Montenegro, H. Bolfarine and V. H. Lachos, Influence diagnostics for a skew extension of the Grubbs measurement error model, Comm. Statist. Simulation Comput. 38 (2009), 667-681.

[22] L. C. Montenegro, V. H. Lachos and H. Bolfarine, Inference for a skew extension of the Grubbs model, Statist. Papers 51 (2010), 701-715.

[23] F. Osorio, G. A. Paula and M. Galea, On estimation and influence diagnostics for the Grubbs' model under heavy-tailed distributions, Comput. Statist. Data Anal. 53 (2009), 1249-1263.

[24] S. Sahinler and D. Topuz, Bootstrap and jackknife resampling algorithms for estimation of regression parameters, Journal of Applied Quantitative Methods 2 (2007), 188-199.

[25] C. B. Zeller, V. H. Lachos and F. V. Labra, Influence diagnostics for Grubbs' model with asymmetric heavy-tailed distributions, Statist. Papers 55 (2014), 671-690.

[26] H. T. Zhu and S. Y. Lee, Local influence for incomplete data models, J. R. Stat. Soc. Ser. B Stat. Methodol. 63 (2001), 111-126.